Minimum Information About a Microarray Experiment – MIAME for Plant Genomics (MIAME/Plant)

DRAFT – Based on MIAME 1.1, MIAME/Tox (2003) and MIAME/Env (2003)

Version of December 13, 2005.

Introduction

MIAME (Minimum Information About a Microarray Experiment, (Brazma et al., 2001)) is a standard that aims at providing a conceptual structure for the core information to be captured from most microarray experiments. MIAME is embedded into a standard microarray data model and exchange format (Microarray Gene Expression Object Model, MAGE-OM; and the corresponding Markup Language, MAGE-ML) as defined by the MGED Society (Microarray Gene Expression Data Society (Causton and Game, 2003), http://www.mged.org/).

Although the MIAME standard has proven very useful so far, especially for the annotation of labeling and hybridization procedures, measurement data, and array design, it has been limited in its ability to capture domain-specific information about experimental design and sample preparation. A working group including scientists from the Swiss Federal Institute of Technology Zurich (ETHZ), the Nottingham Arabidopsis Stock Center (NASC), The Arabidopsis Information Resource (TAIR) and the European Bioinformatics Institute (EBI) has set out to extend the MIAME standard and to establish a list of controlled vocabularies for plant microarray experiments (MIAME/Plant). The MIAME/plant specific extensions to MIAME are included essentially in the “Experiment Package”, “Protocol” and “BioMaterials” objects of the MGED core ontology.

In contrast to MIAME, MIAME/Plant is both a set of guidelines and a number of formal specifications. Nevertheless, the ontologies and specifications used have been reduced to a minimum and retrieved from current standard ontologies. The proposed draft shall serve as a framework for different plant communities to extend it to a level detailed enough to cover most fields of current plant research.

The MIAME/Plant Ontologies and Controlled Vocabularies

This section provides a list of controlled vocabularies for plant microarray experiments complementing those of the general MIAME 1.1 standard (Brazma et al., 2001). It deals essentially with experiment design and samples used (Experiment, Protocol and BioMaterials packages of MAGE-OM). The document contains those descriptions which are not already present in the general MIAME 1.1 standard. Terms and categories are listed within the MIAME 1.1 framework structure (see MIAME 1.1, draft 6, April 1, 2002)

I Array design description

Identical to MIAME 1.1, draft of April 1, 2002, discussed at MGED4

II Experiment description

1. Plant experimental design

This section is common to all hybridization performed on plant microarray experiments and describes pooling of experiments and experimental design.

1) Pooling of experiments[i]

- Number of individuals in each pool
- Whether the samples were pooled before or after extraction ie: multiple samples to create one extract or multiple extracts from multiple samples
- Genotype of each pool if applicable
- Whether all individuals were grown on 1 plate/field or several
- Whether all individuals were grown on the same day

2) Experimental design[ii]

- number of blocks
- layout of blocks - design
    - randomised vs non-randomised

2. Plant samples used, extract preparation and labeling

This section is focused on three main parts:

- Biosource properties
- Biomaterial manipulations
- Extraction method

1) Biosource properties

- **Plant strain or lineGenotype**
  - Germplasm
    - Stock Centre[iii]
    - name
    - accession[iv]
    - (-ecotype)[v]
    - (-mutant / transgenic)
    - (-subspecies)[vi]
    - (-cultivar)[vii]
  - Ecotype[viii]
    - habitat
    - ecotype name
    - location
    - collected when / by whom
  - Mutant[ix]
    - mutagene
    - mutant gene (locus)
    - inheritance[x]
  - Transgenic[xi]
    - gene name (locus)
    - insert type:
      - cDNA
      - genomic construct
      - transposon
      - T-DNA
      - unknown
      - inverted_repeat
    - construct type:
      - activation_tag
      - gene_trap
      - enhancer_trap
      - promoter_trap
      - promoter_reporter
    - protein_fusion
    - unknown
    - simple_insert

- over-expression
- RNAi
- antisense
- promoter_fusion
- cre-lox_recombination
- promoter
- reporter
- selection marker
- vector name (accession number if available)

- **Starting material**[xii]

  - Seed / whole plant
  - Tissue culture
  - Cell culture
  - Protoplasts

  - name
  - publication or source

- **Developmental stage**

  - MIAME/Plant will use the developmental ontology that is being developed by the Plant Ontology Consortium, which includes the Boyes key (http://www.plantontology.org/ontology/index.html) and cereal plant growth stages (GRO, ftp://ftp.gramene.org/pub/gramene/ontology/temporal_stages/temporal_gr_ont)

- **Organism part (tissue)**
  - The plant organs section will embrace both anatomical terminology as well as plant architecture. The latter is related to the plant species being described and must therefore not be necessarily be visible for the end-user entering experiment annotation.
    Plant architecture types could be obtained from:
    http://pais.cirad.fr/project.html

    Plant anatomy ontology should refer to work done by the Plant Ontology Consortium, who tries to integrate the existing taxon-specific ontologies into a common plant structure ontology.
    http://www.plantontology.org/ontology/index.html

    Although this work is still in its infancy, it will be integrated as a standard in the long run.

2) Biomaterial manipulations

- **Growth conditions**[xiii]
    - **Growth substrates**
        - Liquid
            - Hydroponics
                - Constituents
            - Sterilisation?
        - Solid
            - Agar
            - Filter paper
            - Nylon membrane
            - Quartz sand
            - Sterilisation?
        - Natural soil
            - Soil description
            - Soil type
            - Nutrient content
            - pH
            - size distribution
            - organic matter content
        - Commercial soil
            - Soil description
            - Soil type
            - Nutrient content
            - pH
            - size distribution
            - organic matter content
            - Manufacturer(s)
        - Other
            - Aeroponics
                - Constituents
    - **Growth environment**
        - Controlled
            - Greenhouse, with or without artificial light
            - Growth chamber
        - Not-controlled
            - Field
            - Conventional greenhouse
            - Indoor

- **Environmental conditions**

  For ALL growth environments:

  - Light conditions
    - duration day/night (photoperiod)
    - intensity (duration of each if intensity varies)
    - wavelength(s)

  - Light source
    - type
    - cool white fluorescent
    - incandescent
    - laser
    - sun light
    - luminescent
    - high pressure sodium and metal halide
    - manufacturer

  - Watering conditions
  - Relative humidity
    - day
    - night
    - relative

  - Temperature
    - day
    - night
    - average
  - CO2

  - Spacing / density of plants
  - Selection (antibiotic, BASTA…)

  For greenhouses:

  - Plastic covering / sleeves
  - Pots / containers and other accessories
    - Manufacturer
    - Size
  - Aeration - duration and timing

  For cell culture:

  - Media used
    - Publication
    - Media manufacturer
    - Modifications to media

- Strength / concentration of media
- Media pH
- sugar type / content
- Vitamins
- Minerals
- Antibiotic type

For field environments:

- Duration and timing of rainfall
- Climate
  - Vapour pressure deficit (vp)
  - Temperature
  - Relative humidity (rh)
- Irrigation methods
  - border-strip
  - check-basin
  - corrugation
  - flooding
  - furrow
  - sprinkler
  - sub-irrigation
  - wild flooding
  - soil fertility
  - soil tillage

Growth control agents

- Herbicide
- Pesticide
- Fungicide
- Fertiliser
- Nutrients - hormones

Harvesting conditions:

- As environmental (if any changes)

- **Treatment type**

  - **abiotic factors**
    - light
      - wavelength
      - intensity
      - photoperiod
    - temperature
      - heat
      - cold
      - freezing

- mechanical
  - wounding
  - touch
  - wind
- atmospheric
- osmotic or salt
- water
  - anoxia
  - hypoxia
  - drought
- plant nutrient
  - surplus
  - deficiency
  - standard
- types of chemicals
  - heavy metals
  - hormone
  - toxin
  - inhibitor
  - growth regulator
- substrate compaction

- **biotic factors**

  - pathogen or microbe (fungal or bacterial)
    - organism
    - strain
    - dosage
    - incubation conditions
  - animals
    - organism
    - type of effect:
      - wounding
      - grazing
      - feeding (e.g. aphids)
      - mutualistic
  - other plants
    - concurrence
      - nutrients
      - light
      - space
    - shading
    - plant parasites

  - viruses
    - organism
    - strain
    - dosage
    - incubation conditions

- **Stress factor**
    - degree of stress
    - timing of stress (developmental stage)

- **Seed stratification treatment (if experimental treatment)**
    - Temperature
    - Hormonal
    - Duration
    - Moisture / humidity
    - Scarification method used
        - Duration and intensity of treament
        - Chemical
        - Mechanical
        - Brigham and Hoover method
- **Seed sterilisation treatment**
- **Vernalisation treatment**
    - temperature
    - length of vernalization
    - description
    - growth environment
        - cold room
        - growth chamber
        - refridgerator

- **Separation technique**[xiv]

    - Isolation techniques
        - Laser Capture
        - Microdissection
        - Trimming
        - Scarpel

3) Extraction method

- Quantity extracted[xv]

- Extraction source[xvi]
    - Fresh sample
    - Freeze-dried

- Extraction method[xvii]
    - Kit
    - Manufacturer
    - Manual
    - Publication (Author, journal etc) of established method

- Amplification[xviii]
    - PCR
    - Temperature cycles
    - Annealing
    - Primers
    - Time
    - Number of cycles

References

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet **29,** 365-371.

Causton, H.C., and Game, L. (2003). MGED comes of age. Genome Biol **4,** 351.

Craigon, D.J., James, N., Okyere, J., Higgins, J., Jotham, J., and May, S. (2004). NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. Nucleic Acids Res **32 Database issue,** D575-577.

Stoeckert, C.J., Jr., Causton, H.C., and Ball, C.A. (2002). Microarray databases: stanmgeddards and ontologies. Nat Genet **32 Suppl,** 469-473.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L., and Gruissem, W. (2004). GENEVESTIGATOR: Arabidopsis Microarray Database and Analysis Toolbox. Plant Physiology **136**.

[i] Use Pooling protocol, create it as an object in MAGE-ML in the Protocol_package, use parameters objects to create the no of individuals in a pool. Using MAGE-ML pools can be created at various levels; i.e. at the BioSource (i.e. the starting biological material) or BioSample level, or at the (RNA) extract level. If there is a mixed genotype, different biosamples should be created.

[ii] In MGEDOntology --> MGEDCoreOntology --> ExperimentPackage --> ExperimentesignType --> Methodologicaldesign --> Here new instances in the MO might be needed, such as randmised_block_design; block_design; latin_square_design.

[iii] In MGEDOntology --> MGEDCoreOntology --> BioMaterialPackage --> BioMaterialCharacteristics --> BioSourceProvider. Where BioSourceProvider is defined as “The resource (e.g. company, hospital or geographical location) used to obtain or purchase the biomaterial.” In addition, stock centres such as NASC are approved instances of MGEDOntology

--> MGEDCoreOntology --> DescriptionPacakage --> Database --> StrainOrLinedatabase. Other stock centres can be added if required.

[iv] The accession can linked with the property “has_accession” from the database.

[v] Ecotype is MGEDOntology --> MGEDCoreOntology --> BioMaterialPackage --> BioMaterialCharacteristics --> StrainOrLine --> Ecotype, where Ecotype is defined as “a biotype resulting from selection in a particular habitat”.

[vi] The subspecies should be linked to the organism name, and the taxonomy id pertaining to that particular subspecies should be linked from an external organism db such as NCBI taxonomy db.

[vii] Cultivar is MGEDOntology --> MGEDCoreOntology --> BioMaterialPackage --> BioMaterialCharacteristics --> StrainOrLine --> Cultivar in MO.

[viii] Ecotype information regarding location from where collection took place etc should be linked from an external authorized source whenever possible rather then attempting to capture the info in its entirety. If an external db source is not available then one should use the BioSourceProvider class in MO.

[ix] In MGEDOntology --> MGEDCoreOntology --> BioMaterialPackage --> BioMaterialCharacteristics --> GeneticModification, an approved instance of this in induced_mutation which is defined as “the modification of the genetic material (either coding or non-coding) of an organism by mutagenic compounds or irradiation.”

[x] Inheritance may require a new class and/or new instance in MO.

[xi] These are instances in MGEDOntology --> MGEDCoreOntology --> BioSequencePackage --> BioSequenceType . Where the BioSequenceType are defined as “Controlled terms for descriptors indicating the type of biosequence. Types may be physical (e.g. BAC, cDNA clone), or computational (e.g. unigene cluster, consensus).” Some of these instances may need to be added in MO such as vector name.

[xii] The starting material is described in the growth or treatment protocols. The plant organ excised for RNA extraction may be described by the TAIR anatomy ontology. This encompasses terms for cell culture, but may need to be extended as it only includes two terms. The plant ontology may be of use here.

[xiii] These may be mapped to the Plant Environment Ontology available from the OBO site http://obo.sourceforge.net/main.html

[xiv] In MGEDOntology --> MGEDCoreOntology --> BioMaterialPackage --> Action (The action of stabilizing an organism prior to treatment) --> ComplexAction. Here, approved instances cover some of the above terms others can be added if appropriate.

[xv] In MAGE-ML quantity of RNA extracted can be specified as a parameter of the Protocol.

[xvi] Both freeze_dried & fresh_sample are instances of BioSourceType; in MGEDOntology --> MGEDCoreOntology --> BioMaterialPackage --> BiomaterialCharacteristics --> BioSourceType

[xvii] These can be parsed out as parameters attached to the RNA extraction protocol in MAGE-ML.

[xviii] These can be parsed out as parameters in the protocol parameters and is currently done so in ArrayExpress.